Guerra's interpolation using Derrida-Ruelle cascades.

Dmitry Panchenko[*], *Michel Talagrand*[†]

Abstract

New results about Poisson-Dirichlet point processes and Derrida-Ruelle cascades allow us to express Guerra's interpolation entirely in the language of Derrida-Ruelle cascades and to streamline Guerra's computations. Moreover, our approach clarifies the nature of the error terms along the interpolation.

Key words: Sherrington-Kirkpatrick model, Poisson-Dirichlet point process.

Mathematics Subject Classification: 60K35, 82B44

1 Introduction.

The interpolation invented by Francesco Guerra in [3] is one of the most important results in the mathematical theory of the Sherrington-Kirkpatrick model [9]. Guerra showed for the first time in [3] how the Parisi formula [7] appears naturally as an upper bound on the free energy. This was a major step toward the rigorous proof of this formula in [12]. One can define Guerra's interpolation in terms of Derrida-Ruelle cascades [8] similarly to Aizenman-Sims-Starr interpolation [2]; this greatly simplifies the computation leading to the upper bound on the free energy ([1], [2]). However, in order to prove that the upper bound is sharp one needs to understand precisely the error terms along the interpolation as in [12] (see also [6]) and Guerra's original representation is much better suited for this analysis. In this paper we obtain new results about Poisson-Dirichlet point processes and Derrida-Ruelle cascades that allow us to express Guerra's interpolation entirely in the language of the cascades and, in particular, to easily obtain Guerra's representation of the error terms from the corresponding representation via Derrida-Ruelle cascades. This interplay not only streamlines the computations but also helps us understand Guerra's interpolation on the conceptual level.

We consider a Gaussian Hamiltonian $H_N(\boldsymbol{\sigma})$ indexed by spin configurations $\boldsymbol{\sigma} \in \Sigma_N = \{-1,+1\}^N$ with covariance

$$\mathbb{E} H_N(\boldsymbol{\sigma}^1) H_N(\boldsymbol{\sigma}^2) = \xi(R_{1,2}) \tag{1.1}$$

[*]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA and Department of Mathematics, Texas A&M University, College Station, TX, USA. Email: panchenk@math.tamu.edu. Partially supported by NSF grant.

[†]Equipe d'Analyse de l'Institut Mathématique, Paris, France Email: spinglass@talagrand.net.

where

$$R_{1,2} = \frac{1}{N} \boldsymbol{\sigma}^1 \cdot \boldsymbol{\sigma}^2 = \frac{1}{N} \sum_{i \le N} \sigma_i^1 \sigma_i^2$$

is called the overlap of configurations $\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2$ and ξ is a smooth convex function such that $\xi(0) = 0$. Given external field parameter $h \in \mathbb{R}$, free energy is defined by

$$F_N = \frac{1}{N} \mathbb{E} \log \sum_{\boldsymbol{\sigma}} \exp\Bigl(H_N(\boldsymbol{\sigma}) + h \sum_{i \le N} \sigma_i \Bigr). \tag{1.2}$$

The external field term $h \sum \sigma_i$ will play no special role in our considerations so for simplicity of notations it will be omitted.

Guerra's interpolation. Let us first recall Guerra's construction. Given $k \ge 1$, consider sequences $\boldsymbol{m}$ and $\boldsymbol{q}$ such that

$$0 = m_0 < m_1 < \ldots < m_{k-1} < m_k = 1$$

and

$$0 = q_0 < q_1 < \ldots < q_k < q_{k+1} = 1.$$

Consider a matrix

$$Z = (z_{il}) \quad \text{for} \quad 1 \le i \le N \quad \text{and} \quad 0 \le l \le k \tag{1.3}$$

of independent Gaussian r.v. such that $\mathbb{E} z_{il}^2 = \xi'(q_{l+1}) - \xi'(q_l)$, i.e. the coordinates of each column are i.i.d. Let

$$s = (s_1, \ldots, s_N) \quad \text{where} \quad s_i = \sum_{0 \le l \le k} z_{il}.$$

For $0 \le t \le 1$ we define an interpolating Hamiltonian by

$$H_t(\boldsymbol{\sigma}) = \sqrt{t} H_N(\boldsymbol{\sigma}) + \sqrt{1-t}\, s \cdot \boldsymbol{\sigma}. \tag{1.4}$$

Consider $X_k = \log \sum_{\boldsymbol{\sigma}} \exp H_t(\boldsymbol{\sigma})$ and recursively for $1 \le l \le k$ define

$$X_{l-1} = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l X_l \tag{1.5}$$

where $\mathbb{E}_l$ denotes the expectation in (z_{ip}) for $1 \le i \le N$ and $l \le p \le k$. By construction, X_l is a function of (z_{ip}) for $p \le l$. This definition is slightly different from [12], where X_l denoted what we call X_{l-1}, but this indexing will be more convenient when we define Guerra's interpolation in terms of Derrida-Ruelle cascades. Finally, we consider

$$\varphi(t) = N^{-1}\, \mathbb{E} X_0. \tag{1.6}$$

It should be obvious that $\varphi(1) = F_N$ and $\varphi(0)$ can be easily computed since all coordinates decouple and as a result $\varphi(0)$ does not depend on N. Let $\theta(x) = x\xi'(x) - \xi(x)$ and for any $a, b \in \mathbb{R}$ define

$$\Delta(a,b) = \xi(a) - a\xi'(b) + \theta(b). \tag{1.7}$$

By convexity of ξ, $\Delta(a,b) \ge 0$. The following holds.

Theorem 1 (Guerra) *We have,*

$$\varphi'(t) = -\frac{1}{2}\theta(1) + \frac{1}{2}\sum_{1\le r\le k}(m_r - m_{r-1})\theta(q_r) - \frac{1}{2}\sum_{1\le r\le k}(m_r - m_{r-1})\mu_r\big(\Delta(R_{1,2}, q_r)\big), \tag{1.8}$$

where μ_l will be described below.

Definition of μ_r. Fix $1 \le r \le k$. Let

$$W_l = \exp m_l(X_l - X_{l-1}) \quad \text{for} \quad 1 \le l \le k.$$

Notice that by definition of X_l, W_l depends only on (z_{ip}) for $p \le l$. Consider two copies Z^1, Z^2 of Z such that for all $1 \le i \le N$

$$z_{il}^1 = z_{il}^2 \quad \text{for} \quad l \le r-1 \quad \text{and} \quad z_{il}^1, z_{il}^2 \quad \text{are independent for} \quad r \le l. \tag{1.9}$$

This means that the columns 0 through $r-1$ of Z^1, Z^2 are completely correlated and all other columns are independent. We consider Hamiltonians H_t^1 and H_t^2 as above defined in terms of Z^1 and Z^2 correspondingly and define X_l^1, X_l^2 and W_l^1, W_l^2 accordingly. Then, for a function $f : \Sigma_N^2 \to \mathbb{R}$ we define

$$\mu_r(f) = \mathbb{E}\prod_{1\le l<r} W_l^1 \prod_{r\le l\le k} W_l^1 W_l^2 \,\langle f\rangle \tag{1.10}$$

where $\langle\cdot\rangle$ is the Gibbs' average on Σ_N^2 with respect to Hamiltonian

$$H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = H_t^1(\boldsymbol{\sigma}^1) + H_t^2(\boldsymbol{\sigma}^2).$$

Notice that in the first product for $l < r$ we could also write W_l^2 since in this case by construction $W_l^1 = W_l^2$.

Alternative definition of μ_r. Fix $1 \le r \le k$. Consider a sequence $\boldsymbol{n}$ such that

$$n_l = m_l/2 \quad \text{for} \quad l < r \quad \text{and} \quad n_l = m_l \quad \text{for} \quad r \le l. \tag{1.11}$$

In the notations of the first definition let $Y_k = \log\sum_{\boldsymbol{\sigma}^1,\boldsymbol{\sigma}^2}\exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$ and recursively for $1 \le l \le k$ define

$$Y_{l-1} = \frac{1}{n_l}\log \mathbb{E}_l \exp n_l Y_l.$$

Let $V_l = \exp n_l(Y_l - Y_{l-1})$ for $1 \le l \le k$. Then, (1.10) is equivalent to

$$\mu_r(f) = \mathbb{E}\prod_{1\le l\le k} V_l\,\langle f\rangle, \tag{1.12}$$

where again $\langle\cdot\rangle$ denotes the Gibbs' average with respect to the Hamiltonian $H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2)$.

To see that these definitions are the same, it is a simple exercise to show by induction that $V_l = W_l^1 W_l^2$ for $r \le l \le k$ and $V_l = W_l^1 = W_l^2$ for $l < r$ (see Lemma 2.7 in [12]).

Guerra's interpolations via Derrida-Ruelle cascades. We will now define Guerra's interpolation in the language of Derrida-Ruelle cascades similarly to [2].

Given $0 < m < 1$, consider a Poisson point process Π of intensity measure $x^{-1-m}dx$ on $(0,\infty)$. Let $(u_n)_{n\geq 1}$ be a decreasing enumeration of Π and $w_n = u_n / \sum_l u_l$. The distribution of (w_n) is called Poisson-Dirichlet distribution $PD(m,0)$. We will identify a sequence (u_n) with a point process Π and simply call (u_n) itself a Poisson point process.

Let us recall the construction of Derrida-Ruelle cascades (see, for example, [8], [5] or [2]) which involves construction of several processes indexed by $\alpha \in \mathbb{N}^k$. Let us consider a sequence

$$0 < m_1 < m_2 < \ldots < m_k < 1.$$

We start by constructing a family of point processes on the real line as follows.

(i) Let $(u_{n_1})_{n_1\geq 1}$ be a decreasing enumeration of a Poisson point process on $(0,\infty)$ with intensity measure $x^{-1-m_1}dx$.

(ii) Recursively for $2 \leq l \leq k$, for all $(n_1,\ldots,n_{l-1}) \in \mathbb{N}^{l-1}$ we define independent Poisson point processes $(u_{n_1\ldots n_{l-1}n_l})_{n_l\geq 1}$ with intensity measure $x^{-1-m_l}dx$ independent of all previously constructed processes $(u_{n_1\ldots n_j})$ for $j \leq l-1$.

(iii) For $\alpha = (n_1,\ldots,n_k) \in \mathbb{N}^k$ we define $v_\alpha = \prod_{1\leq l\leq k} u_{n_1\ldots n_l}$ and $w_\alpha = v_\alpha \big/ \sum_\alpha v_\alpha$.

The reason why the sum $\sum v_\alpha$ is well defined follows easily from the properties of Poisson point processes (see, for example, [2], [5]). We assume that $m_k < 1$ is because the sum of Poisson point process corresponding to $m_k = 1$ is not well defined (equal to $+\infty$ a.s.). In the interpolation that we will now describe one should formally treat the last step corresponding to $m_k = 1$ differently but this simple modification will unnecessarily complicate the notations. Instead, for simplicity of notations, we will work with $m_k < 1$ and then formally let $m_k \to 1$.

Let $Z = (z_0, z_1, \ldots, z_k)$ be a column representation of a Gaussian matrix in (1.3). Let us define a sequence Z_α of copies of Z as follows.

(i) Let $(z_{n_1})_{n_1\geq 1}$ be i.i.d. copies of z_1.

(ii) Recursively for $2 \leq l \leq k$, for all $(n_1,\ldots,n_{l-1}) \in \mathbb{N}^{l-1}$ we define independent sequences $(z_{n_1\ldots n_{l-1}n_l})_{n_l\geq 1}$ of i.i.d. copies of z_l independent of all $(z_{n_1\ldots n_j})$ for $j \leq l-1$.

(iii) For all $\alpha = (n_1,\ldots,n_k) \in \mathbb{N}^k$ we define $Z_\alpha = (z^\alpha_{il}) = (z_0, z_{n_1}, z_{n_1 n_2}, \ldots, z_{n_1\ldots n_k})$.

Let

$$s^\alpha = (s^\alpha_1, \ldots, s^\alpha_N) \quad \text{where} \quad s^\alpha_i = \sum_{0\leq l\leq k} z^\alpha_{il}.$$

It is easy to check that, by construction, for any $\alpha, \beta \in \mathbb{N}^k$

$$\mathbb{E} s^\alpha_i s^\beta_i = \xi'(q_{\alpha\wedge\beta}) \quad \text{and} \quad \mathbb{E} s^\alpha_i s^\beta_j = 0 \quad \text{for} \quad i \neq j \tag{1.13}$$

where

$$\alpha \wedge \beta = \begin{cases} \min\{l \geq 1 : \alpha_l \neq \beta_l\} & \text{if} \quad \alpha \neq \beta \\ k+1 & \text{if} \quad \alpha = \beta. \end{cases} \tag{1.14}$$

For $0 \leq t \leq 1$ we define a Hamiltonian

$$H_t(\boldsymbol{\sigma}, \alpha) = \sqrt{t} H_N(\boldsymbol{\sigma}) + \sqrt{1-t}\, s^\alpha \cdot \boldsymbol{\sigma} \tag{1.15}$$

and define

$$\varphi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha, \boldsymbol{\sigma}} w_\alpha \exp H_t(\boldsymbol{\sigma}, \alpha). \tag{1.16}$$

Based on the properties of Derrida-Ruelle cascades we will see that $\varphi(t)$ is equal to Guerra's interpolation in (1.6). The definition (1.16) is similar to the Aizenman-Sims-Starr interpolation in [2] with one difference that here we omit an additional term in (1.15). In the present setting, due to the properties of Derrida-Ruelle cascades, adding this extra term is a matter of taste. Not adding this term as the advantage to give an interpolation identical to Guerra's in (1.6). Let us consider a Gibbs probability measure Γ on $\Sigma_N \times \mathbb{N}^k$ defined by

$$\Gamma\big\{(\boldsymbol{\sigma}, \alpha)\big\} \sim w_\alpha \exp H_t(\boldsymbol{\sigma}, \alpha). \tag{1.17}$$

Theorem 2 *We have*

$$\varphi'(t) = -\frac{1}{2}\theta(1) + \frac{1}{2}\mathbb{E}\big\langle \theta(q_{\alpha\wedge\beta})\big\rangle - \frac{1}{2}\mathbb{E}\big\langle \Delta(R_{1,2}, q_{\alpha\wedge\beta})\big\rangle \tag{1.18}$$

where $\langle\cdot\rangle$ *is the Gibbs average with respect to* $\Gamma^{\otimes 2}$.

Proof. By (1.16) and (1.17),

$$\varphi'(t) = \frac{1}{2\sqrt{t}}\mathbb{E}\big\langle H_N(\boldsymbol{\sigma})\big\rangle - \frac{1}{2\sqrt{1-t}}\mathbb{E}\big\langle s^\alpha \cdot \boldsymbol{\sigma}\big\rangle.$$

Using (1.1) and (1.13), Gaussian integration by parts easily implies that this is equal to

$$\begin{aligned} \varphi'(t) &= \frac{1}{2}(\xi(1) - \xi'(1)) - \frac{1}{2}\mathbb{E}\big\langle \xi(R_{1,2}) - R_{1,2}\xi'(q_{\alpha\wedge\beta})\big\rangle \\ &= -\frac{1}{2}\theta(1) + \frac{1}{2}\mathbb{E}\big\langle \theta(q_{\alpha\wedge\beta})\big\rangle - \frac{1}{2}\mathbb{E}\big\langle \Delta(R_{1,2}, q_{\alpha\wedge\beta})\big\rangle. \end{aligned}$$

and this finishes the proof.

□

This proof illustrates that the computation of the derivative in this version of Guerra's interpolation is a simple exercise compared to the original computation of Theorem 1 in [3]. However, in Theorem 1 the corresponding error terms were defined much more precisely and a priori it is not at all obvious how this can be deduced from (1.18). As the following shows, the second term in (1.18) is equal to the second term in (1.8).

Theorem 3 *For all* $1 \leq r \leq k$ *and for all* $0 \leq t \leq 1$,

$$\mathbb{E}\langle I(\alpha\wedge\beta = r)\rangle = \mathbb{E}\Gamma^{\otimes 2}\{\alpha\wedge\beta = r\} = m_r - m_{r-1}. \tag{1.19}$$

This implies that

$$\mathbb{E}\big\langle \theta(q_{\alpha\wedge\beta})\big\rangle = \sum_{1\le r\le k} \mathbb{E}\big\langle I(\alpha\wedge\beta = r)\big\rangle\,\theta(q_r) = \sum_{1\le r\le k}(m_r - m_{r-1})\theta(q_r).$$

It remains to understand the last term in (1.18). Note that in each error term in the last sum in (1.8), the overlap $R_{1,2}$ is compared to a fixed value q_r. Therefore, it seems natural that fixing $\alpha\wedge\beta = r$ in the Gibbs average in (1.18) would produce a corresponding term in (1.8). This turns out to be true but the proof will require new results about Poisson-Dirichlet point processes and Derrida-Ruelle cascades.

Theorem 4 *For $1\le r\le k$, we have*

$$\mathbb{E}\big\langle \Delta(R_{1,2}, q_{\alpha\wedge\beta}) I(\alpha\wedge\beta = r)\big\rangle = (m_r - m_{r-1})\mu_r\big(\Delta(R_{1,2}, q_r)\big). \tag{1.20}$$

The alternative definition of μ_r above played an important role in the proof of Parisi formula in [12] and one might be interested in the corresponding representation via Derrida-Ruelle cascades if one, for example, wishes to write the interpolation in [12] for coupled copies via the cascades. This can be expressed as follows. Let (Z^1, Z^2) be a pair of matrices defined in (1.9). Let $\boldsymbol{n}$ be a sequence defined in (1.11) and let $w_\alpha^{(r)}$ be the Derrida-Ruelle cascades corresponding to parameters given by $\boldsymbol{n}$. Next, we generate a sequence $(Z^1, Z^2)_\alpha$ as above by treating a pair of matrices as a block matrix with twice as many rows. We define a Hamiltonian on $\Sigma_N^2\times\mathbb{N}^k$ by

$$\begin{aligned} H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \alpha) &= \sqrt{t}H_N(\boldsymbol{\sigma}^1) + \sqrt{1-t}\, s^{1,\alpha}\cdot\boldsymbol{\sigma}^1 \\ &+ \sqrt{t}H_N(\boldsymbol{\sigma}^2) + \sqrt{1-t}\, s^{2,\alpha}\cdot\boldsymbol{\sigma}^2 \end{aligned} \tag{1.21}$$

and define a Gibbs' measure Γ_r on $\Sigma_N^2\times\mathbb{N}^k$ by

$$\Gamma_r\big\{(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \alpha)\big\} \sim w_\alpha^{(r)} \exp H_t(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2, \alpha). \tag{1.22}$$

The following holds.

Theorem 5 *For any function $f:\Sigma_N^2\to\mathbb{R}$ we have $\mu_r(f) = \mathbb{E}\langle f\rangle_r$ and, in particular,*

$$\mu_r\big(\Delta(R_{1,2}, q_r)\big) = \mathbb{E}\langle \Delta(R_{1,2}, q_r)\rangle_r \tag{1.23}$$

where $\langle\cdot\rangle_r$ is the average with respect to the Gibbs measure Γ_r in (1.22).

2 Properties of Poisson-Dirichlet point processes.

In this section we obtain new results regarding the Poisson-Dirichlet point process and in the next section we will generalize them to Derrida-Ruelle cascades. These results will immediately imply Theorems 3, 4 and 5. First, let us state a well-known property of Poisson-Dirichlet point process (see [8] or Lemma 6.5.15 in [10]).

Lemma 1 *Let $0 < m < 1$. If (u_n) is a Poisson point process with intensity measure*

$$d\mu = x^{-1-m}dx \quad on \quad (0, \infty)$$

and $U_n > 0$ are i.i.d. random variables such that $\mathbb{E}U_n < \infty$ then

$$(u_n U_n) \quad and \quad (u_n (\mathbb{E}U_1^m)^{1/m})$$

are both Poisson point processes with the same intensity measure $\mathbb{E}U_1^m d\mu$.

Next, we will prove a result that contains the main idea of the paper. Let $\mathcal{X}$ be a complete separable metric space that we will also view as a measurable space with Borel σ-algebra. Consider an i.i.d. sequence (X_n, Y_n) with distribution ν on $\mathbb{R} \times \mathcal{X}$ independent of (u_n) and such that $X_n > 0$. Let ν_1, ν_2 denote the marginals of ν and ν_x denote a regular conditional distribution of Y given $X = x$. Suppose that $\mathbb{E}X < \infty$ and define by ν_m a probability measure on $\mathcal{X}$

$$\nu_m(B) = \int \frac{x^m}{\mathbb{E}X^m} \nu_x(B) d\nu_1(x)$$

which is obviously a distribution of Y under the change of density $X^m/\mathbb{E}X^m$, i.e. for any measurable function ϕ,

$$\int \phi(y) d\nu_m(y) = \frac{\mathbb{E}X^m \phi(Y)}{\mathbb{E}X^m}.$$

The following holds.

Lemma 2 *Poisson point process $(u_n X_n, Y_n)$ has the same distribution as a point process $((\mathbb{E}X^m)^{1/m} u_n, Y_n')$ where (Y_n') is an i.i.d. sequence independent of (u_n) with distribution ν_m.*

Proof. By the marking theorem ([4]) a point process (u_n, X_n, Y_n) is a Poisson point process with intensity measure $\mu \otimes \nu$ on $(0, \infty) \times (0, \infty) \times \mathcal{X}$. By the mapping theorem ([4]), $(u_n X_n, Y_n)$ is a Poisson point process with intensity measure given by the image of $\mu \otimes \nu$ under the mapping $(u, x, y) \to (ux, y)$ if this measure has no atoms. Let us compute this image measure. Given two measurable sets $A \subseteq (0, \infty)$ and $B \subseteq \mathcal{X}$,

$$\mu \otimes \nu(ux \in A, y \in B) = \int \mu(u : ux \in A) \nu_x(B) d\nu_1(x).$$

For $x > 0$ we have

$$\mu(u : xu \in A) = \int I(xu \in A) x^{-1-m} dx = u^m \int I(z \in A) z^{-1-m} dz = u^m \mu(A)$$

and, therefore,

$$\mu \otimes \nu(ux \in A, y \in B) = \int x^m \mu(A) \nu_x(B) d\nu_1(x) = \mathbb{E}X^m \mu(A) \otimes \nu_m(B).$$

Since measure $\mathbb{E}X^m \mu$ is the intensity measure of a Poisson point process $((\mathbb{E}X^m)^{1/m} u_n)$ this finishes the proof.

□

As an application of Lemma 2 we will give a new simple proof of Theorem 6.4.5 in [10].

Corollary 1 *If (X_n, Y_n) are i.i.d. such that $X \geq 1$ and $\mathbb{E}X^2, \mathbb{E}Y^2 < \infty$ then*

$$\mathbb{E}\frac{\sum u_n Y_n}{\sum u_n X_n} = \frac{\mathbb{E}X^{m-1}Y}{\mathbb{E}X^m}, \tag{2.1}$$

$$\mathbb{E}\frac{\sum u_n^2 Y_n^2}{(\sum u_n X_n)^2} = (1-m)\frac{\mathbb{E}X^{m-2}Y^2}{\mathbb{E}X^m}, \tag{2.2}$$

$$\mathbb{E}\frac{\sum_{n\neq m} u_n u_m Y_n Y_m}{(\sum u_n X_n)^2} = m\Big(\frac{\mathbb{E}X^{m-1}Y}{\mathbb{E}X^m}\Big)^2. \tag{2.3}$$

Proof. If we denote by $c = (\mathbb{E}X^m)^{1/m}$ then by Lemma 2,

$$\mathbb{E}\frac{\sum u_n Y_n}{\sum u_n X_n} = \mathbb{E}\frac{\sum (u_n X_n)(Y_n/X_n)}{\sum u_n X_n} = \mathbb{E}\frac{\sum (u_n c)(Y_n/X_n)'}{\sum u_n c} = \mathbb{E}\frac{X^m}{\mathbb{E}X^m}\frac{Y}{X}$$

since the markings $(Y_n/X_n)'$ are independent of (u_n) and the distribution is given by the change of density $X^m/\mathbb{E}X^m$. Similarly,

$$\begin{aligned}\mathbb{E}\frac{\sum u_n^2 Y_n^2}{(\sum u_n X_n)^2} &= \mathbb{E}\frac{\sum (u_n X_n)^2 (Y_n/X_n)^2}{(\sum u_n X_n)^2} \\ &= \mathbb{E}\frac{\sum (u_n c)^2 (Y_n/X_n)'^2}{(\sum u_n c)^2} = \mathbb{E}\frac{X^m}{\mathbb{E}X^m}\frac{Y^2}{X^2}\,\mathbb{E}\frac{\sum u_n^2}{(\sum u_n)^2}.\end{aligned}$$

To finish the proof of (2.2) it remains to use a well-known fact (Corollary 2.2 in [8] or Proposition 1.2.7 in [10])

$$\mathbb{E}\sum w_n^2 = (1-m). \tag{2.4}$$

Finally,

$$\begin{aligned}\mathbb{E}\frac{\sum_{n\neq m} u_n u_m Y_n Y_m}{(\sum u_n X_n)^2} &= \mathbb{E}\frac{\sum_{n\neq m} (u_n X_n)(u_m X_m)(Y_n/X_n)(Y_m/X_m)}{(\sum u_n X_n)^2} \\ &= \mathbb{E}\frac{\sum_{n\neq m} (u_n c)(u_m c)(Y_n/X_n)'(Y_m/X_m)'}{(\sum u_n c)^2} = m\Big(\mathbb{E}\frac{X^m}{\mathbb{E}X^m}\frac{Y}{X}\Big)^2\end{aligned}$$

since by (2.4), $\mathbb{E}\sum_{n\neq m} w_n w_m = 1 - \mathbb{E}\sum w_n^2 = m$.

□

3 Properties of Derrida-Ruelle cascades.

Let us construct a general random process Z_α indexed by $\alpha \in \mathbb{N}^k$ in a much more general way than the random matrix process in the second version of Guerra's interpolation above. Consider complete separable metric spaces $\mathcal{X}_1, \ldots, \mathcal{X}_k$ which we also view as measurable spaces with Borel σ-algebras and for $1 \leq l \leq k$ let

$$\mathcal{X}^l = \mathcal{X}_1 \times \ldots \times \mathcal{X}_l.$$

Consider a probability measure ν on $\mathcal{X}_1$ and for $1 \le l < k$ consider regular conditional distributions

$$\nu_l(\cdot|x) \quad \text{on} \quad \mathcal{X}_{l+1} \quad \text{for} \quad x \in \mathcal{X}^l. \tag{3.1}$$

We generate a process

$$Z_\alpha = (z_{n_1}, z_{n_1 n_2}, \ldots, z_{n_1 n_2 \ldots n_k}) \in \mathcal{X}^k$$

according to the following recursive procedure.

(i) Generate i.i.d. random variables $(z_{n_1})_{n_1 \ge 1}$ with distribution ν.

(ii) Recursively over $2 \le l \le k$, given $(z_{n_1}, \ldots, z_{n_1 \ldots n_{l-1}})$ for all $n_1 \ldots n_{l-1} \in \mathbb{N}$ we generate i.i.d. sequences $(z_{n_1 \ldots n_{l-1} n_l})_{n_l \ge 1}$ with distributions

$$\nu_l(\cdot|z_{n_1}, \ldots, z_{n_1 \ldots n_{l-1}}) \tag{3.2}$$

independently for all $n_1, \ldots, n_{l-1}$.

(iii) For each $\alpha = (n_1, \ldots, n_k) \in \mathbb{N}^k$ we define $Z_\alpha = (z_{n_1}, z_{n_1 n_2}, \ldots, z_{n_1 \ldots n_k}) \in \mathcal{X}^k$.

For convenience of notations, given $\alpha = (n_1, \ldots, n_k)$ we denote for $1 \le l \le k$,

$$\alpha^l = (n_1 \ldots n_l), \quad u_{\alpha^l} = u_{n_1 \ldots n_l} \quad \text{and} \quad v_{\alpha^l} = \prod_{1 \le j \le l} u_{\alpha^j} \tag{3.3}$$

so that $v_{\alpha^{l+1}} = v_{\alpha^l} u_{\alpha^l}$. Given $Z_\alpha \in \mathcal{X}^k$ we denote

$$z_{\alpha^l} = z_{n_1 \ldots n_l} \quad \text{and} \quad Z_{\alpha^l} = (z_{n_1}, \ldots, z_{n_1 \ldots n_l}).$$

Consider a measurable function $X : \mathcal{X}^k \to \mathbb{R}$ such that $\mathbb{E} \exp X(Z_\alpha) < \infty$. Let $X_\alpha = X(Z_\alpha)$ and recursively for $1 \le l \le k$ define

$$X_{\alpha^{l-1}} = \frac{1}{m_l} \log \mathbb{E}_l \exp m_l X_{\alpha^l} \tag{3.4}$$

where $\mathbb{E}_l$ denotes the expectation conditionally on $(Z_{\alpha^{l-1}})_{\alpha \in \mathbb{N}^k}$ and

$$W_{\alpha^l} = \exp m_l (X_{\alpha^l} - X_{\alpha^{l-1}}). \tag{3.5}$$

Thus, both X_{α^l} and W_{α^l} are functions of Z_{α^l}. In particular, $X_0 := X_{\alpha^0}$ is a constant. It will be convenient to think of W_{α^l} as a function of two variables

$$W_{\alpha^l} = W_l(Z_{\alpha^{l-1}}, z_{\alpha^l}).$$

Let us now generate another process Z'_α exactly the same way as Z_α with one modification that instead of (3.2) the distribution of $(z'_{n_1 \ldots n_{l-1} n_l})_{n_l \ge 1}$ conditionally on $Z'_{\alpha^{l-1}} = (z'_{n_1}, \ldots, z'_{n_1 \ldots n_{l-1}})$ will be given by

$$W_l(Z'_{\alpha^{l-1}}, x)\, d\nu_l(x|Z'_{\alpha^{l-1}}). \tag{3.6}$$

This is a probability measure because by (3.4), (3.5) and (3.2),

$$\int W_l(Z'_{\alpha^{l-1}}, x)\, d\nu_l(x|Z'_{\alpha^{l-1}}) = \mathbb{E}_l \exp m_l(X_{\alpha^l} - X_{\alpha^{l-1}}) = 1.$$

For $1 \le l \le k$, let us define

$$e_{\alpha^l} = \exp(X_{\alpha^l} - X_{\alpha^{l-1}}). \tag{3.7}$$

The following in the generalization of Lemma 2.

Lemma 3 *The point processes*

$$(u_{\alpha^1} e_{\alpha^1}, \ldots, u_{\alpha^k} e_{\alpha^k}, Z_{\alpha^k}) \quad \textit{and} \quad (u_{\alpha^1}, \ldots, u_{\alpha^k}, Z'_{\alpha^k}) \tag{3.8}$$

on $\mathbb{R}^{+k} \times \mathcal{X}^k$ *have the same distribution.*

Proof. The proof is by induction on k. The case $k = 1$ immediately follows from Lemma 2. Consider $k > 1$. By induction assumption, point processes

$$(u_{\alpha^1} e_{\alpha^1}, \ldots, u_{\alpha^{k-1}} e_{\alpha^{k-1}}, Z_{\alpha^{k-1}}) \quad \text{and} \quad (u_{\alpha^1}, \ldots, u_{\alpha^{k-1}}, Z'_{\alpha^{k-1}}) \tag{3.9}$$

have the same distribution. If we write

$$Z_{\alpha^k} = (Z_{\alpha^{k-1}}, z_{\alpha^k}) \quad \text{and} \quad Z'_{\alpha^k} = (Z'_{\alpha^{k-1}}, z'_{\alpha^k})$$

it suffices to show that conditionally on the processes (3.9), the two processes

$$\big(u_{\alpha^k} e_{\alpha^k}, z_{\alpha^k}\big) \quad \text{and} \quad \big(u_{\alpha^k}, z'_{\alpha^k}\big) \tag{3.10}$$

have the same distribution . Let us write $\alpha^k = (\alpha^{k-1}, n)$ and for a fixed α^{k-1} look at the point process $\big(u_{\alpha^k} e_{\alpha^k}, z_{\alpha^k}\big)_{n\ge 1}$. Let us apply Lemma 2 to this sequence conditionally on (3.9). By (3.4),

$$\mathbb{E}_k e_{\alpha^k}^{m_k} = \mathbb{E}_k \exp m_k(X_{(\alpha^{k-1},n)} - X_{\alpha^{k-1}}) = 1$$

and, therefore, by Lemma 2, the point processes

$$\big(u_{\alpha^k} e_{\alpha^k}, z_{\alpha^k}\big)_{n\ge 1} \quad \text{and} \quad \big(u_{(\alpha^{k-1},n)}, z'_{(\alpha^{k-1},n)}\big)_{n\ge 1} \tag{3.11}$$

have the same distribution, where $z'_{(\alpha^{k-1},n)}$ is distributed as $z_{(\alpha^{k-1},n)}$ under the change of density

$$\frac{e_{\alpha^k}^{m_k}}{\mathbb{E}_k e_{\alpha^k}^{m_k}} = \exp m_k(X_{(\alpha^{k-1},n)} - X_{\alpha^{k-1}}) = W_k(Z_{\alpha^{k-1}}, z_{\alpha^k}).$$

By construction, $z_{(\alpha^{k-1},n)}$ are distributed according to $\nu_l(\cdot|Z_{\alpha^{k-1}})$ and the change of density defines a distribution

$$W_k(Z_{\alpha^{k-1}}, x)\, d\nu_k(x|Z_{\alpha^{k-1}})$$

which is precisely the distribution (3.6) for $l = k$. Since conditionally on (3.9) processes (3.11) are generated independently for all α^{k-1}, this shows that conditionally on (3.9) both

processes in (3.10) are generated according to the same distribution and this finishes the proof.

□

In particular, Lemma 3 implies that the processes

$$v_\alpha \exp(X_\alpha - X_0) = \prod_{1\le l\le k} u_{\alpha^l} e_{\alpha^l} \quad \text{and} \quad v_\alpha = \prod_{1\le l\le k} u_{\alpha^l} \tag{3.12}$$

have the same distribution, which generalizes Theorem 5.4 in [2]. As a consequence we get (Proposition 2 in [5])

$$\mathbb{E} \log \sum w_\alpha \exp X_\alpha = X_0. \tag{3.13}$$

Using (3.13) one only needs to compare the definitions to observe the equality of (1.6) and (1.16). Using (3.12), Lemma 3 also implies that

$$\big(v_\alpha \exp(X_\alpha - X_0), Z_\alpha\big) \quad \text{and} \quad \big(v_\alpha, Z'_\alpha\big) \tag{3.14}$$

have the same distribution. As we will now show, this immediately implies Theorems 4 and 5. Moreover, the change of density (3.6) makes the definition of measures μ_r in Guerra's interpolation in (1.8) much more transparent.

In addition to X, consider a measurable function $Y : \mathcal{X}^k \to \mathbb{R}$ such that $\mathbb{E}Y^2(Z_\alpha) < \infty$ and let $Y_\alpha = Y(Z_\alpha)$. Theorem 5 is an immediate consequence of the following.

Theorem 6 *We have*

$$\mathbb{E} \frac{\sum_\alpha v_\alpha (\exp X_\alpha)\, Y_\alpha}{\sum_\alpha v_\alpha \exp X_\alpha} = \mathbb{E} \prod_{1\le l\le k} W_{\alpha^l} Y_\alpha. \tag{3.15}$$

Proof. The proof follows immediately by (3.14), because

$$\begin{aligned} \mathbb{E} \frac{\sum_\alpha v_\alpha (\exp X_\alpha)\, Y(Z_\alpha)}{\sum_\alpha v_\alpha \exp X_\alpha} &= \mathbb{E} \sum w_\alpha Y(Z'_\alpha) \\ &= \mathbb{E} Y(Z'_\alpha) = \mathbb{E} \prod_{1\le l\le k} W_{\alpha^l} Y_\alpha, \end{aligned}$$

where in the second line α is fixed and the last equality holds since the distribution of Z'_α is defined by the change of density (3.6).

□

Let us now fix $1 \le r \le k$. Consider a measurable function $Y : \mathcal{X}^k \times \mathcal{X}^k \to \mathbb{R}$ such that $\mathbb{E}Y^2(Z_\alpha, Z_\beta) < \infty$ for any $\alpha, \beta \in \mathbb{N}^k$ and let $Y_{\alpha,\beta} = Y(Z_\alpha, Z_\beta)$. Let us consider fixed $\alpha, \beta \in \mathbb{N}^k$ such that $\alpha \wedge \beta = r$. Let

$$M_r = \mathbb{E} \prod_{l<r} W_{\alpha^l} \prod_{l\ge r} W_{\alpha^l} W_{\beta^l} Y_{\alpha,\beta}.$$

Clearly, M_r depends on α and β only through $r = \alpha \wedge \beta$. Theorem 4 is an immediate consequence of the following.

Theorem 7 *We have*

$$\mathbb{E}\frac{\sum_{\alpha\wedge\beta=r} v_\alpha v_\beta \exp(X_\alpha+X_\beta)\, Y_{\alpha,\beta}}{\bigl(\sum_\alpha v_\alpha \exp X_\alpha\bigr)^2} = (m_r - m_{r-1})M_r. \tag{3.16}$$

Proof. Again, by (3.14)

$$\mathbb{E}\frac{\sum_{\alpha\wedge\beta=r} v_\alpha v_\beta \exp(X_\alpha+X_\beta)\, Y_{\alpha,\beta}}{\bigl(\sum_\alpha v_\alpha \exp X_\alpha\bigr)^2} = \mathbb{E}Y(Z_\alpha', Z_\beta')\,\mathbb{E}\sum_{\alpha\wedge\beta=r} w_\alpha w_\beta,$$

where $\mathbb{E}Y(Z_\alpha', Z_\beta')$ is taken for any fixed α and β such that $\alpha\wedge\beta=r$. By construction, this expectation is equal to M_r because the distribution of Z_α' is defined by the change of density (3.6) and, because, since $\alpha\wedge\beta=r$, the function $Y(Z_\alpha', Z_\beta')$ depends on one copy $z_{\alpha^l}'=z_{\beta^l}'$ for $l<r$ and on two independent copies z_{α^l}' and z_{β^l}' for $l\geq r$. It remains to show that

$$\mathbb{E}\sum_{\alpha\wedge\beta=r} w_\alpha w_\beta = m_r - m_{r-1}. \tag{3.17}$$

Given $\alpha\in\mathbb{N}^k$ let us write $\alpha^r=(a,n)$ for $a\in\mathbb{N}^{r-1}$ and $n\in\mathbb{N}$. If $\alpha\wedge\beta=r$ then $\beta^r=(a,m)$ for $m\neq n$. In the notations of (3.3) let us define $U_{(a,n)}=\sum_{\gamma:\gamma^r=(a,n)}\prod_{r<l\leq k} u_{\gamma^l}$. Then

$$\sum_{\alpha\wedge\beta=r} w_\alpha w_\beta = \frac{\sum_{\alpha\wedge\beta=r} v_\alpha v_\beta}{\bigl(\sum_\alpha v_\alpha\bigr)^2} = \frac{\sum_a v_a^2 \sum_{n\neq m}(u_{(a,n)}U_{(a,n)})(u_{(a,m)}U_{(a,m)})}{\bigl(\sum_a v_a\sum_n u_{(a,n)}U_{(a,n)}\bigr)^2}.$$

A sequence $(U_{(a,n)})$ is i.i.d. by construction and, therefore, by Lemma 1, a point process $(u_{(a,n)}U_{(a,n)})$ has the same distribution as $(u_{(a,n)}c)$ where $c=(\mathbb{E}U_{(a,n)}^{m_r})^{1/m_r}<\infty$. As a result,

$$\mathbb{E}\sum_{\alpha\wedge\beta=r} w_\alpha w_\beta = \mathbb{E}\frac{\sum_a v_a^2\sum_{n\neq m} u_{(a,n)}u_{(a,m)}}{\bigl(\sum_a v_a\sum_n u_{(a,n)}\bigr)^2}.$$

Using that

$$\sum_{n\neq m} u_{(a,n)}u_{(a,m)} = \Bigl(\sum_n u_{(a,n)}\Bigr)^2 - \sum_n u_{(a,n)}^2 = U_a^2 - \sum_n u_{(a,n)}^2$$

where we introduced $U_a=\sum_n u_{(a,n)}$, we can write

$$\mathbb{E}\frac{\sum_a v_a^2\sum_{n\neq m} u_{(a,n)}u_{(a,m)}}{\bigl(\sum_a v_a\sum_n u_{(a,n)}\bigr)^2} = \mathbb{E}\frac{\sum_a (v_aU_a)^2}{\bigl(\sum_a v_aU_a\bigr)^2} - \mathbb{E}\frac{\sum_{a,n} v_{(a,n)}^2}{\bigl(\sum_{a,n} v_{(a,n)}\bigr)^2}. \tag{3.18}$$

By Corollary 3.3 in [8], the process $(v_{a,n}/\sum v_{a,n})$ has Poisson-Dirichlet distribution $PD(m_r,0)$. By Lemma 3 above, the process $(v_aU_a/\sum v_aU_a)$ has the same distribution as the process $(v_a/\sum v_a)$ which again, by Corollary 3.3 in [8], is $PD(m_{r-1},0)$. Therefore, using (2.4) twice implies that the right hand side of (3.18) is equal to $(1-m_r)-(1-m_{r-1})=m_r-m_{r-1}$. This finishes the proof.

□

Finally, we prove Theorem 3.

Proof of Theorem 3. Let Γ_1 be a marginal on $\mathbb{N}^k$ of measure Γ defined in (1.17). Then

$$\Gamma_1\{\alpha\} = v_\alpha f_\alpha \Big/ \sum_\alpha v_\alpha f_\alpha \quad \text{where} \quad f_\alpha = \sum_{\boldsymbol{\sigma}} \exp H_t(\boldsymbol{\sigma}, \alpha). \tag{3.19}$$

By Lemma 3, conditionally on $H_N(\boldsymbol{\sigma})$ and $(z_{i0}, y_{i0})_{1\leq i\leq N}$, the sequence $(\Gamma_1\{\alpha\})_{\alpha\in\mathbb{N}^k}$ is equal in distribution to the sequence $(w_\alpha)_{\alpha\in\mathbb{N}^k}$ and, consequently, the same is true unconditionally. Therefore,

$$\mathbb{E}\Gamma^{\otimes 2}\{\alpha\wedge\beta = r\} = \mathbb{E}\Gamma_1^{\otimes 2}\{\alpha\wedge\beta = r\} = \mathbb{E}\sum_{\alpha\wedge\beta=r} w_\alpha w_\beta = m_r - m_{r-1},$$

using (3.17). This finishes the proof.

□

References

[1] Aizenman, M., Sims, R., Starr, S. (2003) An extended variational principle for the SK spin-glass model. *Phys. Rev. B*, **68**, 214403.

[2] Aizenman, M., Sims, R., Starr, S. (2006) Mean-Field Spin Glass models from the Cavity-ROSt Perspective. Preprint.

[3] Guerra, F. (2003) Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233**, no. 1, 1-12.

[4] Kingman, J. F. C. (1993) *Poisson processes.* Oxford University Press, New York.

[5] Panchenko, D., Talagrand, M. (2004) Bounds for diluted mean-fields spin glass models. *Probab. Theory Related Fields* **130**, no. 3, 319-336.

[6] Panchenko, D. (2005). Free energy in the generalized Sherrington-Kirkpatrick mean field model. *Rev. Math. Phys.* **17**, no. 7, 793-857.

[7] Parisi, G. (1980) A sequence of approximate solutions to the S-K model for spin glasses. *J. Phys.* **A 13**, L-115.

[8] Ruelle, D. (1987) A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Phys.* **108**, no. 2, 225 - 239.

[9] Sherrington, D., Kirkpatrick, S. (1972) Solvable model of a spin glass. *Phys. Rev. Lett.* **35**, 1792-1796.

[10] Talagrand, M. (2003) Spin Glasses: a Challenge for Mathematicians. Springer-Verlag.

[11] Talagrand, M. (2003) On Guerra's broken replica-symmetry bound. *C. R. Math. Acad. Sci. Paris* **337**, no. 7, 477-480.

[12] Talagrand, M. (2006) Parisi formula. *Ann. of Math. (2)* **163**, no. 1, 221-263.